

Beribis lietuvių kalbos pasaulis skaitmeninių išteklių sistemoje „E. kalba“

Jurgita JAROSLAVIENĖ, Rita MILIŪNAITĖ
Lietuvių kalbos institutas

Pasakojimą apie lietuvių kalbos skaitmeninius išteklius ir jų naudą norime pradėti gera žinia, kad 2020 m. spalio 13 d. Lietuvos Respublikos Seimas sutartinai patvirtino lietuvių kalbai ir jos ateičiai svarbų dokumentą – Lietuvių kalbos plėtros skaitmeninėje terpėje ir kalbos technologijų pažangos 2021–2027 metų gaires. Jas parengė Valstybinės lietuvių kalbos komisijos sudaryta darbo grupė, kurioje dirbo ir šio straipsnio autorės.

Lietuvių kalba, kaip ir daugelis kitų informacinės pasaulio bendruomenės vartojamų kalbų, skaitmeninėje terpėje gyvuoja jau ne pirmą dešimtmetį. Dar prieš keletą metų Lietuva galėjo didžiuotis greičiausiu interneto ryšiu pasaulyje, dabar ir toliau sėkmingai plėtoja informacines technologijas. Jos vis plačiau diegiamos įvairiose veiklos srityse, tokiose kaip valstybės administravimas ir teismai, krašto apsauga, elektroninė bankininkystė, sveikatos apsauga, švietimas, mokslas, kultūra ir jos paveldo saugojimas, energetika, verslas, viešasis transportas, gamtosauga, žiniasklaida ir kt. Koronaviruso pandemijos metu ypač

išryškėjo nuotolinio mokymo ir skaitmeninio jo turinio svarba.

Kur šuoliuoja kalbos technologijos?

Pastarieji keletas metų rodo, kad pasaulis pereina į kokybiškai naują informacinių technologijų kūrimo etapą, pagrįstą dirbtiniu intelektu ir didžiųjų duomenų analize. Vis daugiau kalbama apie daiktų internetą, o išmanieji įrenginiai vis labiau skverbiasi į mūsų gyvenimą ir ima pakeisti rankų darbą.

Svarbi dirbtinio intelekto technologijų dalis yra kalbos technologijos, nes dauguma informacijos, kurią dirbtiniam intelektui reikia apdoroti, perteikiama kalbos ženklais. Kad kompiuteriai gebėtų analizuoti sakytinę ir rašytinę kalbą ir tai naudoti įvairioms paslaugoms (pavyzdžiui, mašininiam vertimui, kompiuterinių įrenginių valdymui balsu ir pan.), jiems pirmiausia reikia didelių skaitmeninių kalbos duomenų sankaupų. Lietuvių kalba ir čia neturi atsilikti nuo pasaulyje nuolat tobulinamų kalbos technologijų.

Minėtos Gairės turi padėti užtikrinti visavertį lietuvių kalbos vartojimą skaitmeninėje terpėje, įtvirtinti ir palaikyti lietuvių kalbos statusą informacinėje visuomenėje. Tam reikia gausinti skaitmeninius kalbos išteklius – tekstynus ir garsynus, plėtoti kalbos technologijas ir jų pagrindu kurti viešąsias paslaugas, kad nė viena visuomenės grupė ar regionas nejaustų skaitmeninės atskirties, o kitakalbiai galėtų lengviau integruotis į Lietuvos visuomenę. Kalbos technologijos turi padėti stiprinti Lietuvos visuomenės ir išeivijos ryšius, mažinti ir lietuviškai kalbančios bendruomenės atskirtį globalioje žinių visuomenėje.

Kas yra „E. kalba“?

Ne vienas žmogus iš savo patirties patvirtintų, kad tikrąją gimtosios kalbos vertę galima pajusti tik pagyvenus kitų kalbų apsuptyje. Kai savõsios kalbõs ribos susitraukia ir ji paprastai gyvuoja tik nedideliame artimųjų ar tautiečių bendruomenės būrelyje, tada išryškėja šios kalbos savumas, lengvumas, malonus skambėjimas, apima jos ilgesys. Tai ir yra gimtosios kalbos jausmas.

Kaip šį jausmą išsaugoti, kaip neprarasti gimtosios kalbos įgūdžių globalizacijos sąlygomis? Kaip paskatinti vaikus, kad išeivijoje jie greta pagrindinės kalbos norėtų mokytis ir tos, kuri prigimtinė jų tėvams?

Lietuvių kalbos institute baigiama kurti internetinė nemokama skaitmeninių

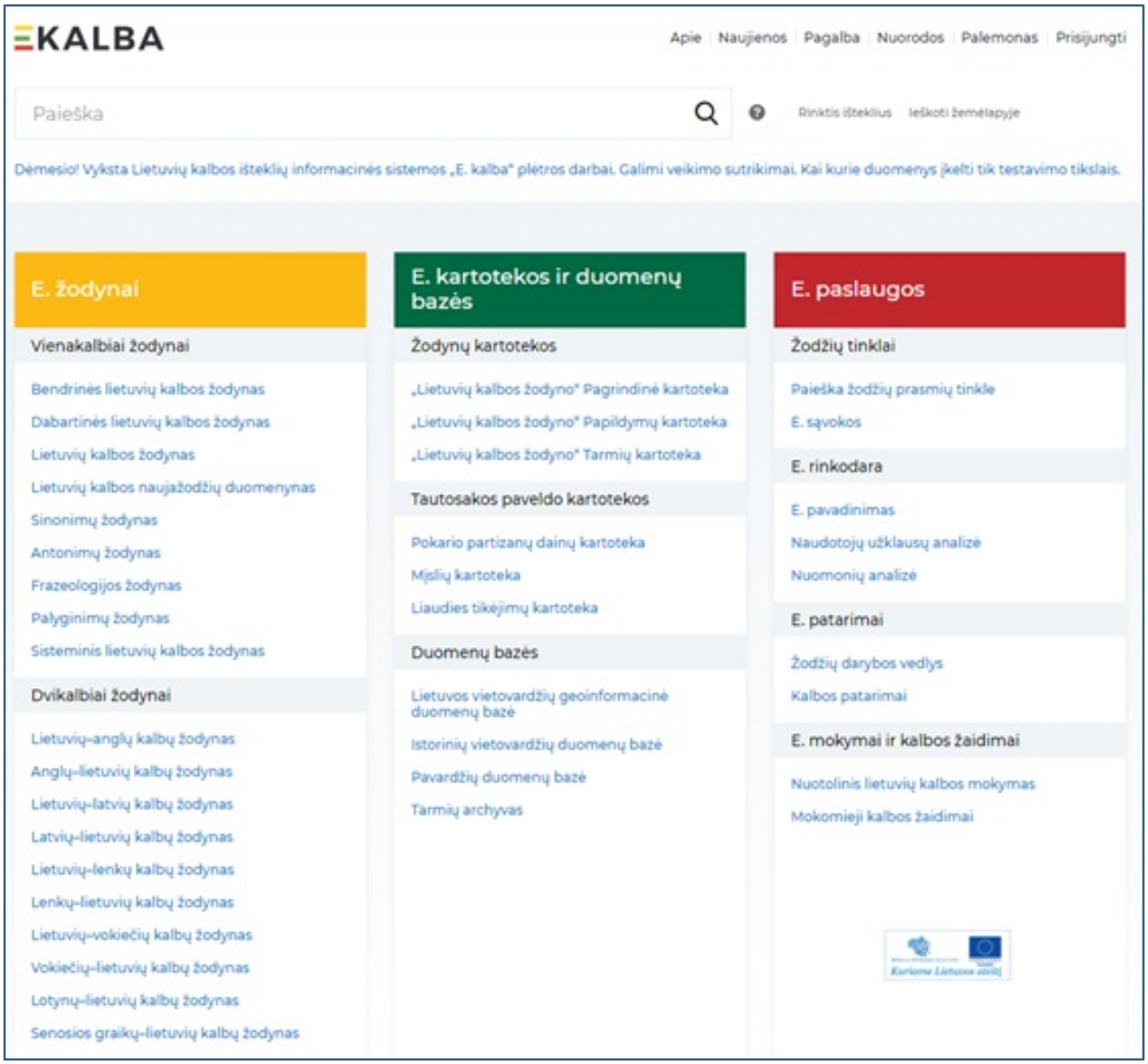
kalbos išteklių sistema „E. kalba“ (šiuo metu prieinama bandomoji „E. kalbos“ sistemos versija, žr. 1 pav.). Tai atnaujinta ir gerokai naujais ištekliais bei paslaugomis papildyta buvusi lietuvių kalbos išteklių sistema LKIIS. Šie ištekliai svarbūs ir kalbos technologijoms plėtoti, ir kasdienėms lietuvių kalbos vartotojų reikmėms.

Ko iš šios išteklių sistemos gali tikėtis jos svetainės lankytojai?

- Išsamios informacijos – daugybės įvairių lietuvių kalbos duomenų vienoje vietoje;
- patikimumo – tai Lietuvių kalbos instituto mokslininkų su partneriais iš Vilniaus universiteto ir Lietuvių kalbos ir literatūros instituto sukurtas ir nuolat atnaujinamas kalbos išteklių lobynas;
- patogumo – vieno langelio principu galima gauti išsamią informaciją apie norimą žodį ar posakį iš įvairių šaltinių;
- naujovių – modernių kalbos paslaugų, skirtų vertėjams, redaktoriams, žiniasklaidininkams, įvairių tekstų kūrėjams, mokytojams ir dėstytojams, mokiniams ir studentams, verslui, taip pat kalbos ir kitų sričių mokslininkams, kalbos technologijų ir skaitmeninių paslaugų kūrėjams, o visiems, kurie nori tobulinti lietuvių kalbos įgūdžius, – kalbos žaidimų.

„E. kalbos“ išteklių sistema sudaryta iš trijų stambių dalių: „E. žodynai“, „E. kartotekos ir duomenų bazės“ ir „E. paslaugos“.

Apie visa tai – plačiau.



1 pav. Skaitmeninių išteklių sistemos „E. kalba“ bandomoji versija: pradžios puslapis

Žodynai

Ši „E. kalbos“ dalis apima 9-is vienakalbius ir 10 dvikalbių žodynų. Prie LKIIS sistemoje jau buvusių vienakalbių žodynų („Dabartinės lietuvių kalbos žodyno“, sinonimų, antonimų, frazeologijos, palyginimų ir sisteminio lietuvių kalbos žodyno) pridėti dar 3 itin svarbūs žodynai, kuriuos iki šiol internete buvo galima rasti tik atskirose svetainėse:

1) norminamasis „Bendrinės lietuvių kalbos žodynas“ – jame profesionaliai aprašoma dabartinė bendrinės lietuvių kalbos leksika; žodynas dar nėra baigtas, bet jis dalimis nuolat skelbiamas internete ir papildomas, o prieš tai svarstomas

Valstybinėje lietuvių kalbos komisijoje;

2) didysis „Lietuvių kalbos žodynas“ – šimtmetį kelių kalbininkų kartų rašytas 20-ies tomų lietuvių kalbos lobynas, į kurį sudėta leksika iš daugiau kaip 1000 sakytinių ir rašytinių šaltinių – nuo pirmųjų lietuviškų raštų iki 20 a. pabaigos, iš tautosakos, tarmių, grožinės bei mokslinės literatūros, periodinės spaudos;

3) naujausius lietuvių kalbos leksikos pokyčius atspindintis Lietuvių kalbos naujažodžių duomenynas, kuris nuolat atnaujinamas renkant informaciją iš interneto žiniasklaidos, socialinių tinklų, lietuvių ir verstinės grožinės literatūros bei kitų šaltinių, taip pat sulaukia daug lankytojų siūlymų įtraukti į duomenyną jų aptiktų naujų žodžių.

Dvikalbiai žodynai ypač reikalingi vertėjams, taip pat visiems, kurie mokosi kalbų. „E. kalbos“ išteklių sistemoje galima naudotis lietuvių–anglų ir anglų–lietuvių, lietuvių–latvių ir latvių–lietuvių, lietuvių–lenkų ir lenkų–lietuvių, lietuvių–vokiečių ir vokiečių–lietuvių, taip pat lotynų–lietuvių ir senosios graikų–lietuvių kalbų žodynais.

Taigi žodynų dalyje pagal įvairius poreikius galima rasti išsamios informacijos apie lietuvių kalbos žodžių, posakių vartojimą, jų atitikmenis kitose kalbose, taip pat kitų minėtų kalbų atitikmenis lietuvių kalboje.

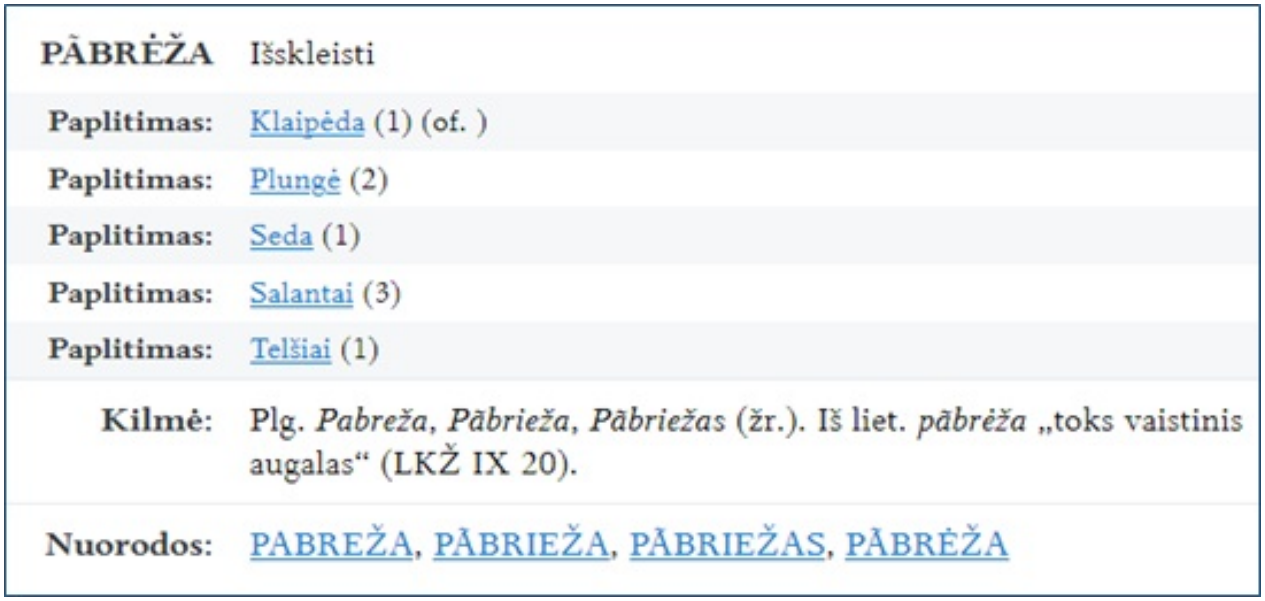
Kartotekos ir kitos duomenų sankaupos

Turbūt esate girdėję, kad minėtas 20-ies tomų „Lietuvių kalbos žodynas“ (LKŽ) parašytas remiantis didžiule – apie 4,5 milijono kortelių turinčia kartoteka. Ją 1902 m. pradėjo kaupti šio žodyno pradininkas kalbininkas Kazimieras Būga, ir per šimtą metų ją rinko ne tik kalbininkai, bet ir mokytojai, kraštotyrininkai, įvairūs kiti talkininkai. Dėl tokios apimties ne viskas galėjo patekti į popierinį žodyną, tačiau ranka įvairiomis rašysenomis išmarginti lapeliai yra neįkainojamas mūsų tautos turtas. Jis ne tik atskleidžia lietuvių kalbos grožį ir turtingumą, bet ir liudija žmonių darbštumą, atsidavimą autentiškų kalbos faktų išsaugojimui. Kartoteka dar tik skenuojama, bet dalį duomenų jau galima pamatyti skaitmeninėje „E. kalbos“ išteklių sistemoje. Duomenys renkami ir toliau, tam skirtos dvi kitos – LKŽ papildymų ir tarmių kartotekos.

Tiems, kurie domisi istorija ir lietuvių kalbos paveldu, turėtų būti įdomios trys tautosakos paveldo kartotekos: Partizanų dainų kartoteka, Mįslių kartoteka ir

Liaudies tikėjimų kartoteka, kurios suskaitmenintos iš Lietuvių literatūros ir tautosakos instituto fondų.

Duomenų bazių skyriuje rasite prieigas prie keturių skaitmeninių išteklių. Neabejotinai bus įdomu paieškoti savo ar giminių pavardės bei sužinoti apie jų kilmę Pavardžių duomenų bazėje (žr. 2 pav.).



PĀBRĖŽA	Išskleisti
Paplitimas:	Klaipėda (1) (of.)
Paplitimas:	Plungė (2)
Paplitimas:	Seda (1)
Paplitimas:	Salantai (3)
Paplitimas:	Telšiai (1)
Kilmė:	Plg. *Pabreža*, *Pābrieža*, *Pābriežas* (žr.). Iš liet. *pābrėža* „toks vaistinis augalas“ (LKŽ IX 20).
Nuorodos:	PABREŽA, PĀBRIEŽA, PĀBRIEŽAS, PĀBRĖŽA

2 pav. Duomenys apie pavardę *Pabrėža* skaitmeninių išteklių sistemoje „E. kalba“

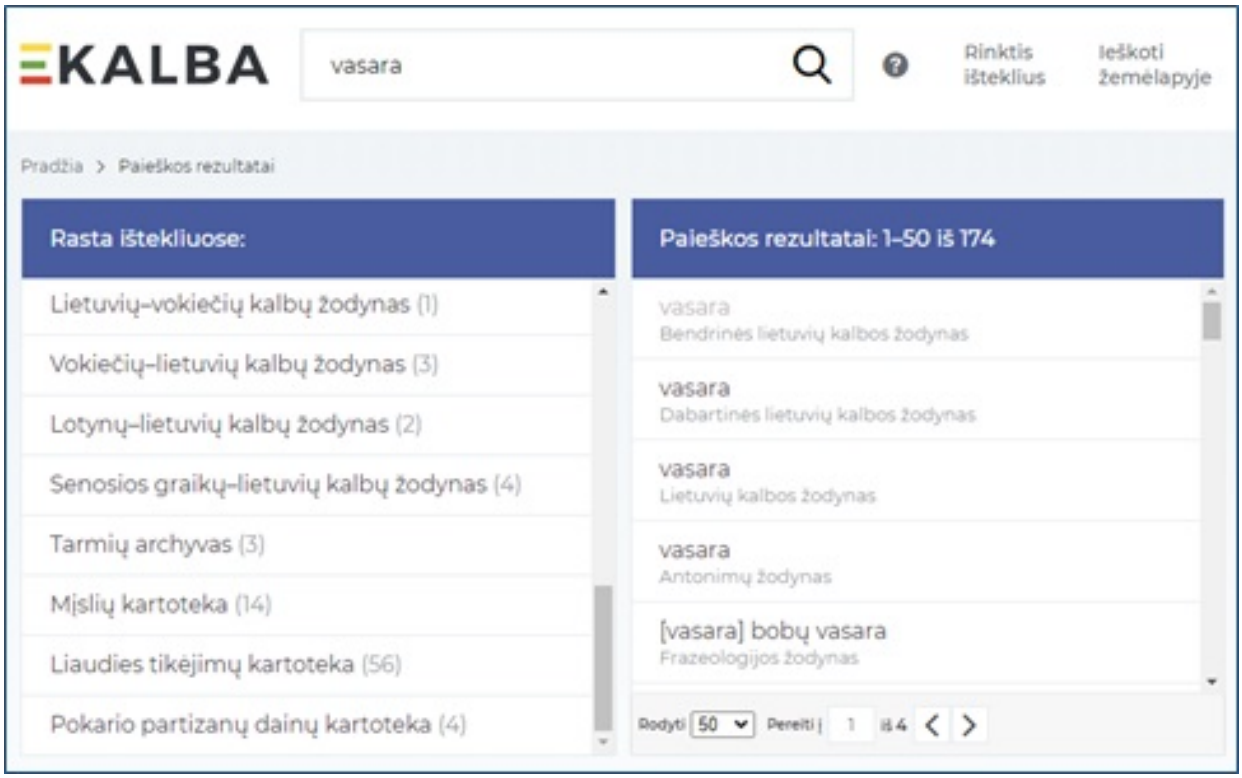
Daug informacijos apie Lietuvos vietų vardus kaupiama Lietuvos vietovardžių geoinformacinėje duomenų bazėje, ji nuolat pildoma.

Istorijos tyrėjams pravarti Istorinių vietovardžių duomenų bazė, kurioje galima sužinoti, kaip per laiką keitėsi kurios nors vietovės pavadinimai, užrašyti iš įvairių istorijos šaltinių.

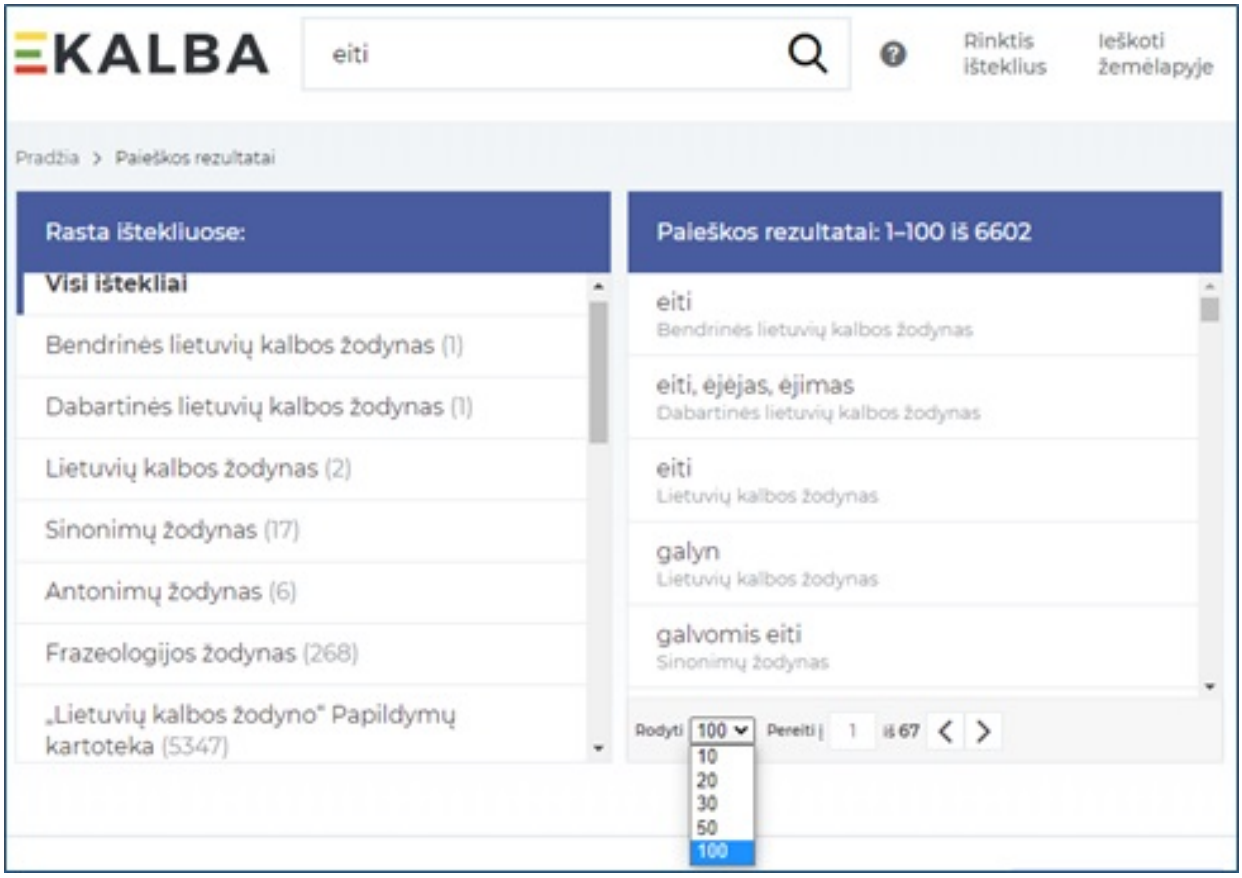
Paieška „E. kalbos“ išteklių sistemoje

Paieška visuose žodynuose ir kartotekose bei duomenų bazėse veikia vieno langelio principu, t. y. norimo žodžio ar frazės galima ieškoti visuose žodynuose iškart. Tarkim, paieškos langelyje surinkę žodį „vasara“, gausime 174 paieškos rezultatus iš 22 išteklių, surinkę žodį „namai“, gausime 627 paieškos rezultatus iš 24 išteklių, surinkę „gerai“, gausime 889 rezultatus iš 19 šaltinių, o surinkę „eiti“ – net 6602 paieškos rezultatus iš daugiau nei 20 šaltinių (plg. 3 ir 4 pav.). Kai „E. kalbos“ sistema bus baigta kurti (planuojama baigti 2021 m. pavasarį), rezultatų su minėtais žodžiais, ko gero, rasime ir dar daugiau, nes atsiras galimybė ieškoti

ir kituose – šiuo metu kuriamuose –ištekliuose, pavyzdžiui, elektroniniuose kalbos patarimuose.



3 pav. Žodžio „vasara“ paieškos rezultatų fragmentas



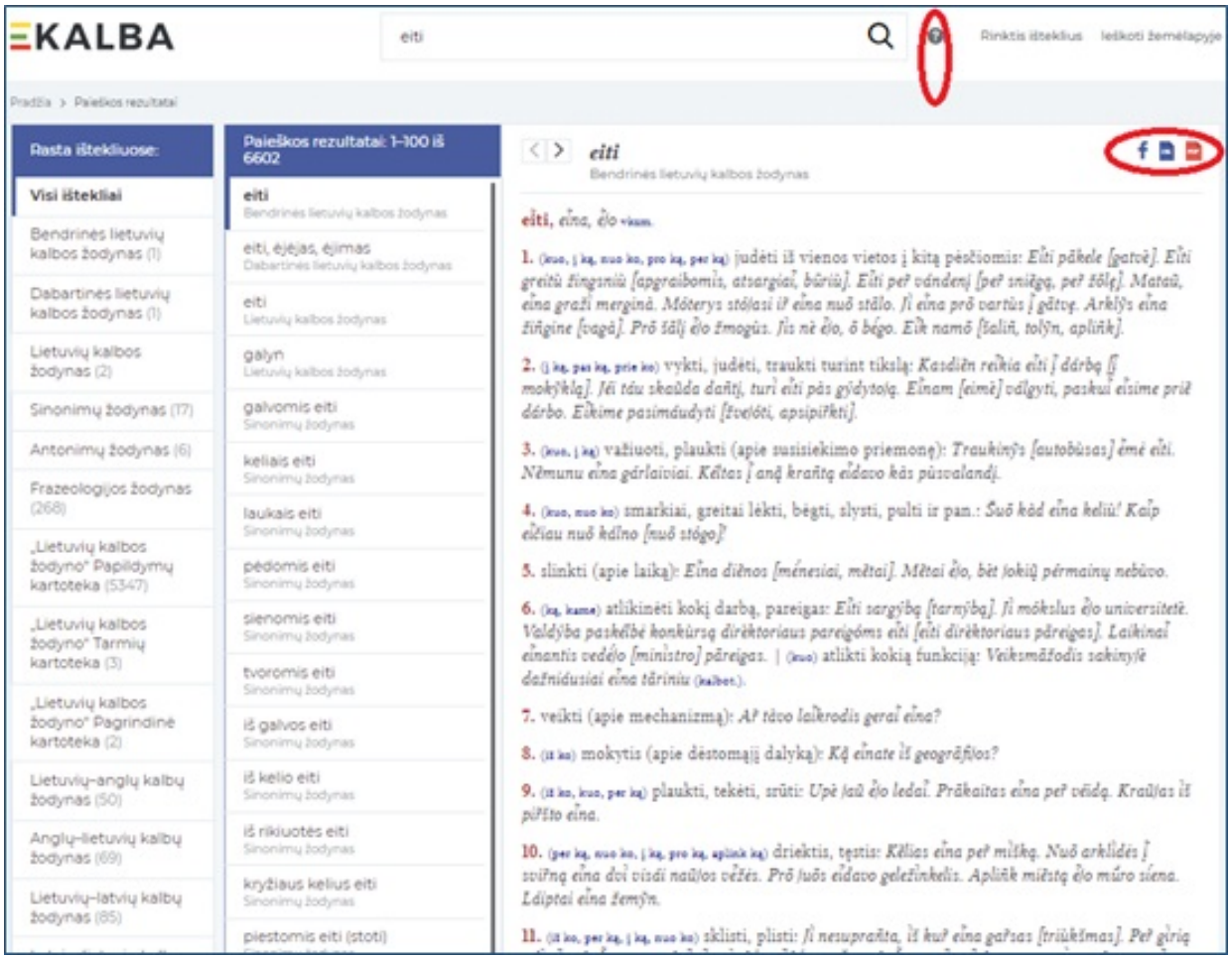
4 pav. Žodžio „eiti“ paieškos rezultatų fragmentas

Toliau iš gautų sąrašų galima rinktis arba rūpimą išteklių, arba konkretų paieškos rezultatą, kurį, beje, galima dar ir parsisiųsti pasirinktu formatu arba pasidalyti feisbuke (plg. 5 pav.). Ieškant rūpimų duomenų, verta pasinaudoti pagalba (žr. apibrauktą klaustuką 5 paveiksle), kur paaiškinta, kaip ieškoti žodžių ar frazių naudojant žvaigždutę (*) ir klaustuką (?). Pavyzdžiui, vienas ar daugiau klaustukų pravers, kai nežinome, kokia to ar tų klaustukų vietoje gali būti raidė arba kai savo klaviatūroje neturime lietuviškų raidžių su diakritikais ą, ę, į, ų, ū, ė, č, š, ž.

Tarkim, įrašę su klaustukais „a?i?“, gausime rezultatų su žodžiais „ačiū, akis, ašis, apie…“; įrašę „sl??ti“, gausime rezultatų su žodžiais „slėgti, slėpti, slysti, slopti, slūgti…“; įrašę „??uolas“, gausime „ąžuolas“, o įrašę „?uolas“ – „suolas, kuolas, buolas, duolas, muolas…“ ir t. t. Kaip matome, kiekvienas klaustukas atitinka po vieną raidę.

Žvaigždutė pravers ieškant daugiau įvairių įvesto žodžio ar fragmento variantų. Tarkim, į paieškos langelį su žvaigždute įrašę „ači*“, gausime rezultatų su žodžiais „ačiū, ačiavimas, ačiulis, ačiuotė, ačiuotojas, ačiuotuvės…“; į paiešką įrašę „*iena“, gausime rezultatų sąrašą su žodžiais „iena, bulviena, burokiena, diena, dobiliena, elniena, ėriena, grybiena, grikiena, higiena, jautiena, kalakutiena, laikraštiena, ląsteliena, laukiniena, naujiena, siena…“; įrašę „šalt*is“, gausime „šaltis, šaltadienis, šaltalankis, šaltėsis, šaltinis, šaltkrepšis, šaltmetis…“.

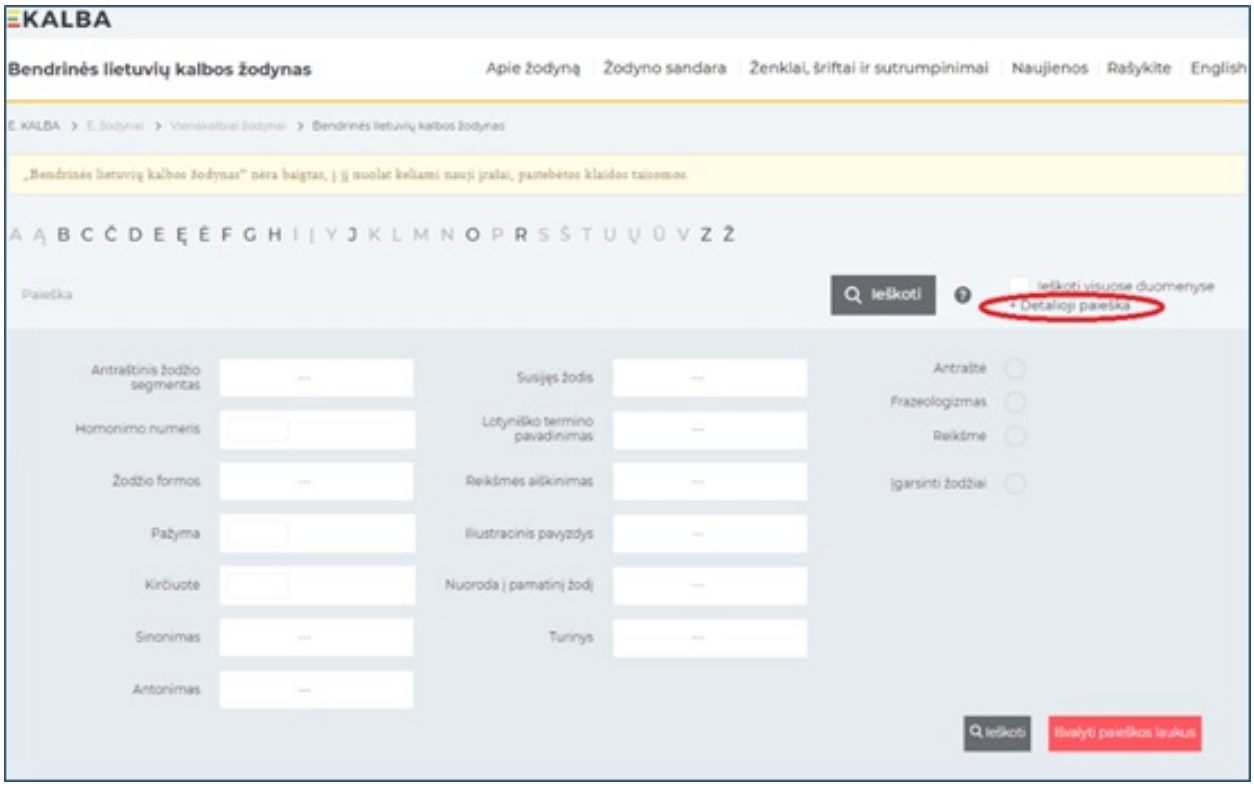
Labai patogu, kad vieno langelio principu rezultatų galima ieškoti viename, visuose arba keliuose pasirinktuose ištekliuose.



5 pav. Iš sąrašo pasirinkto paieškos rezultato fragmentas

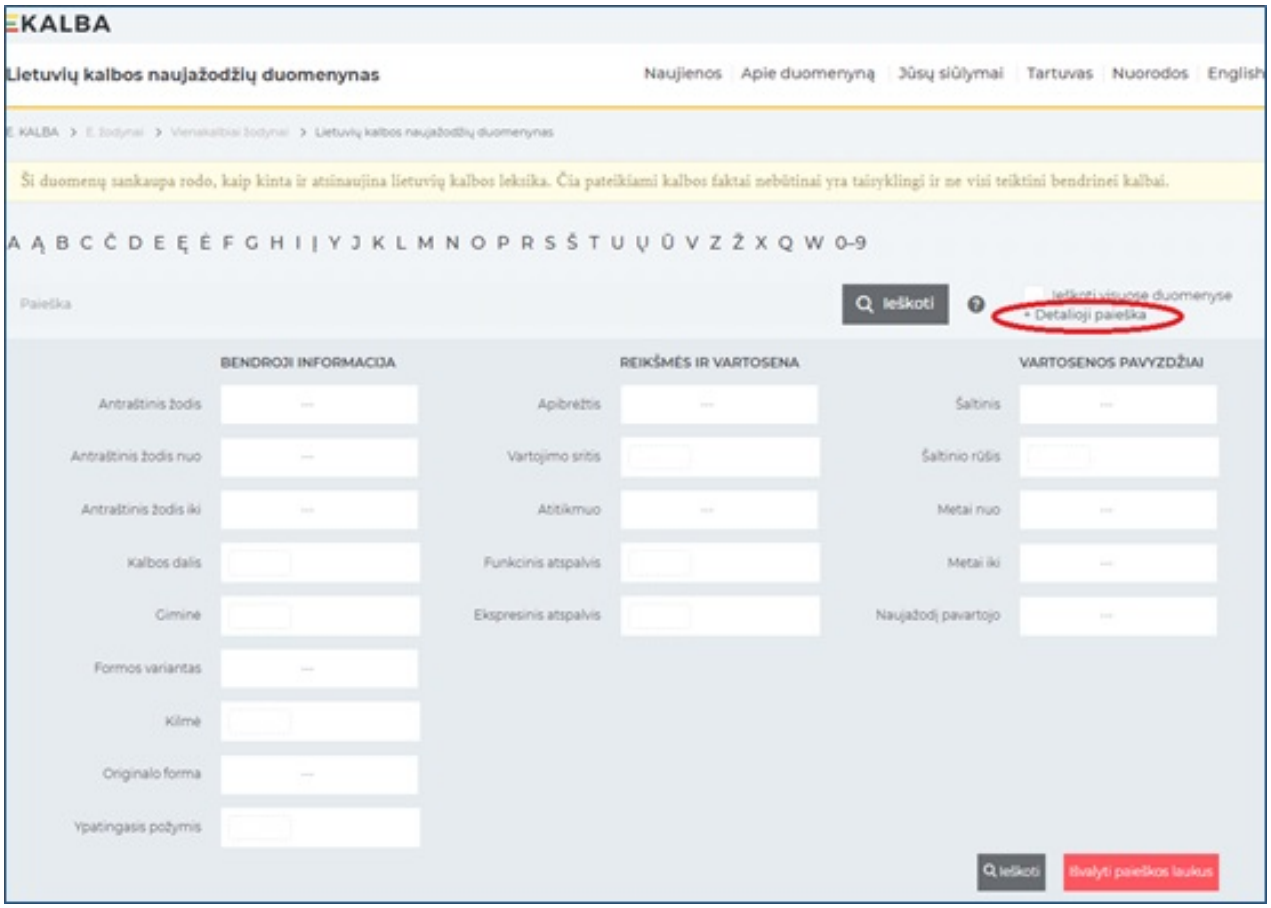
Greitosios paieškos rezultatai prioritetizuojami, kad būtų kuo patogiau ir funkcionaliau: ieškant visuose ištekliuose pirmiausia žodžių ar rūpimų duomenų pateikiama iš naujausių norminių žodynų (plg. 5 pav.).

Daugelis „E. kalbos“ sistemos išteklių turi ir detaliosios paieškos galimybę: pasirinkus konkretų vienakalbį ar dvikalbį žodyną ar kitą šaltinį, galima jame naršyti pagal tam tikrus kriterijus, kuriuos reikėtų pasirinkti atsidarius to šaltinio detaliosios paieškos langą (6 ir 7 paveiksluose apibraukta, ką paspausti, kad atsidarytų detaliosios paieškos langas). Pavyzdžiui, pagal detaliąją paiešką pasirinkę žodžių ar kitų duomenų ieškoti „Bendrinės lietuvių kalbos žodyne“, galime rinktis, ar ieškosime antraštinio žodžio segmento, ar ir žodžio formos, kirčiuotės, sinonimo, antonimo, susijusio žodžio, reikšmės aiškinimo, iliustracinio pavyzdžio, įgarsintų žodžių ir pan. Kadangi šis norminamasis žodynas dar rašomas, jame nebaigti aprašyti žodžiai blyškesni (6 paveiksle matyti, kad nebaigtų aprašyti žodžių abėcėlės raidės taip pat blyškesnės).



6 pav. „Bendrinės lietuvių kalbos žodyno“ detaliosios paieškos langas

Atsidarę Lietuvių kalbos naujažodžių duomenyno detaliosios paieškos langą, naujų žodžių ieškoti galime ir pagal bendrąją informaciją (antraštinį žodį, kalbos dalį, giminę, formos variantą, kilmę, originalo formą, ypatingąjį požymį…), ir pagal reikšmę bei vartoseną (apibrėžtį, vartojimo sritį, atitikmenį, funkcinį ir ekspresinį atspalvį…), ir pagal vartosenos pavyzdžius (šaltinį ir jo rūšį, metus, naujažodį viešai pavartojusių asmenų vardus ir pavardes…). Pavyzdžiui, nežinodami, kokių lietuviškų atitikmenų realiai vartojama vietoj skolinio *aftepartis*, lange „Originalo forma“ surinkę anglišką žodį *afterparty*, pamatysime, kad vartosenoje fiksuoti ir duomenyne aprašyti atitikmenys *dūzgės, povakaris* ir *pratęstuvės*.



7 pav. Naujažodžių duomenyno detaliosios paieškos langas

Greitoji ir detalioji paieška galima ne tik žodynuose ir kartotekose bei duomenų bazėse, bet ir kai kuriose e. paslaugose.

E. paslaugos

Daugiausia naujovių – „E. kalbos“ sistemos dalyje „E. paslaugos“. Iš viso bus galima naudotis 9-iomis paslaugomis, iš kurių net 7-ios naujos.

Šiuo metu kuriamas lietuvių kalbos žodžių prasmių tinklas (ŽT) ir paieška jame. ŽT – tai skaitmeninis išteklius, kuriame žodžiai susisteminami ir tarpusavyje sujungiami įvairiais prasminiais ryšiais. Į „E. kalbos“ ŽT duomenys integruojami ir ryšiai formuojami tarp žodžių iš šių „E. kalbos“ išteklių: „Bendrinės lietuvių kalbos žodyno“, „Dabartinės lietuvių kalbos žodyno“, 20-ies tomų „Lietuvių kalbos žodyno“, Lietuvių kalbos naujažodžių duomenyno, sinonimų, antonimų ir frazeologijos žodynų. ŽT bus siejamas ir su pasaulyje pripažintu anglų kalbos žodžių tinklu, su kuriuo siejami ir kitų pasaulio kalbų žodžių tinklai.

Paslauga „E. sąvokos“ leis gauti informacijos apie dominančią sąvoką iš 3-jų

galimų sričių (ontologijų): medicinos (žmogaus anatomijos terminų), finansų (ekonomikos terminų) ir informacinių technologijų (kompiuterinės technikos ir jos dalių). Žodžių prasmių tinkle ir ontologijose bus galima greitoji ir detalioji paieška.

E. paslauga „Nuomonių analizė“ – tai įrankis, kuris pateiktus naujienų ar socialinių tinklų komentarus, internetinius pokalbius, atsiliepimus ar kitokį tekstą automatiškai įvertins pagal teigiamą, neutralų ir neigiamą atspalvį. Nuomonių iš tekstų analizatorius leis įmonėms išsiaiškinti savo klientų požiūrį į jų prekės ženklą, gaminamus produktus, teikiamas paslaugas, nustatyti visuomenės nuomonę apie asmenis ar įvykius ir kt. Gautus rezultatus vartotojai galės rūšiuoti, parsisiųsti.

Naudotojų užklausų analizė skirta pateiktą užklausą papildyti kalbine informacija: paieškos žodžiui ar frazei bus parinkta prasminiais ryšiais susijusių žodžių.

Paslauga „E. pavadinimas“ padės susikurti lietuvišką įmonės, gaminio, paslaugos ar kitą rūpimą pavadinimą. Įrašius žodį ar žodžius, kuo tiksliau apibūdinančius rūpimą veiklą, gaminį, paslaugą, sritį ar kt., ir paspaudus mygtuką „Ieškoti“, bus pasiūlyta įprastų, retesnių, naujų ar netikėtų žodžių ar žodžių junginių, įvairiais prasminiais ryšiais susijusių su pateikta užklausa. Jeigu yra, bus pasiūlyta ir susijusių tradicinių lietuvių kalbos frazeologizmų. Pasižvalgius po darybos priemones ir pavyzdžius, pavadinimą bus galima susikurti ir naudojantis „Žodžių darybos vedliu“, kuris yra paslaugos „E. patarimai“ dalis,.

„E. patarimai“ susideda iš dviejų e. paslaugų: minėto „Žodžių darybos vedlio“ ir skaitmeninių „Kalbos patarimų“. Šios paslaugos leis produktyviau naudotis lietuvių kalbos galimybėmis, susirasti reikiamą kalbos reiškinį (pavyzdžiui, nenorminių žodžių atitikmenis), kurti naujus žodžius, naudojant lietuvių kalbos žodžių darybos priemones, remiantis žodžių darybos taisyklėmis, gauti išsamią informaciją apie bendrinės kalbos normas ir jų svyravimus. Čia atsiras ir suskaitmenintos prieš keliolika metų pradėtos leisti „Kalbos patarimų“ serijos knygos, ir keletas ankstesnių kalbos rekomendacijų šaltinių. Vartotojai patys galės spręsti, kokias raiškos priemones rinktis.

Norintys pasimokyti lietuvių kalbos nuotoliniu būdu, skaitmeninių išteklių sistemoje „E. kalba“ ras ir nuotolinio mokymo priemonių: 10 pamokų su autentiškais tekstais, vaizdo medžiaga, pratybomis, žodynu ir gramatikos

užduotimis. Taip pat bus ir mokomųjų kalbos žaidimų, skatinančių savarankiškai įtvirtinti lietuvių kalbos rašybos įgūdžius, gausinti raiškos priemonių atsargas. Apie šiuos žaidimus „Pasaulio lietuvyje“ jau buvo pasakota.

Tikimės, kad lietuvių kalbos skaitmeninių išteklių sistema „E. kalba“ taps tokiu interneto tinklalapiu, kurį mielai lankysite ir jausitės savi beribiame lietuvių kalbos pasaulyje.



Straipsnių ciklas „Lietuvių kalbos naujovės ir įdomybės“. Straipsnį rėmė Valstybinė lietuvių kalbos komisija.
Jei norėtumėte publikuoti visą straipsnį ar jo dalį, prašome nurodyti informacijos šaltinį ir autorių.